Mechanisms, Explanation and Understanding in Physics

Dennis Dieks
History and Philosophy of Science
Utrecht University
d.dieks@uu.nl

Abstract

The Scientific Revolution is often associated with a transition to a "mechanistic" world view. However, "mechanization" is not the term that best captures the distinctive nature of modern physics: "mathematization" would be a better characterization. Modern physics attempts to find mathematical relations between quantities, and does not require that these relations be interpreted in terms of mechanisms. Moreover, in modern physics there are cases in which it is unnatural to give the mathematical formalism a mechanistic interpretation, even if "mechanistic" is broadly construed. Both on the level of ontology and that of explanation physics turns out to be more general and liberal than what is suggested by the catchphrase that physics explains by identifying mechanisms. Although mechanistic explanation remains an important conceptual tool, in particular for achieving understanding, it is not the only one available and cannot lay claim to fundamentality.

1 Introduction

In his book "The Mechanization of the World Picture" [1], the historian of science E.J. Dijksterhuis famously described the transition from ancient and medieval to modern science, in particular physics, as the replacement of occult qualities by clear and empirically accessible "mechanical" concepts like the size, velocity and acceleration of particles. Newtonian mechanics, the culmination of the Scientific Revolution, established mathematically formulated laws between quantities of this sort. The term "mechanization" seems apt for this transition, and it is true that the treatment of physical problems on the basis of mechanics became an ideal of physical science in the seventeenth, eighteenth and nineteenth centuries. However, as Dijksterhuis

himself already noted in the epilogue of his book [1, p. 498], one should bear in mind that

> the science called mechanics had emancipated itself in the 17^{th} century from its origins in the study of machines, and had developed into an independent branch of mathematical physics, dealing with the motion of material objects and finding in the theory of machines only one of its numerous practical applications.

As Dijksterhuis makes clear in his epilogue, with hindsight the fundamental contrast between ancient and medieval physics on the one hand and (early) modern physics on the other is not that the former sometimes uses explanations involving teleology or analogies with organisms whereas the latter models processes with concepts that come from the world of machines. Rather, the basic difference is that while ancient and medieval physics occasionally used mathematical tools, modern physics is essentially mathematical, defining core concepts in a mathematical way and formulating laws in mathematical language. Mathematical reasoning is by its nature abstract, and it is not self-evident that the metaphor of a "machine" will always be natural or even applicable for mathematically described processes, not even for processes within the domain of the science of mechanics itself.

The novel mathematical frameworks that were invented for the treatment of mechanical problems in the eighteenth and nineteenth centuries strengthen this point. As we shall see, these new approaches have led in the direction of growing abstraction and have made thinking in terms of what we intuitively would call "mechanisms" less than obvious. The development of new branches of physics in the nineteenth and twentieth centuries have further contributed to the general picture of increasing abstraction and distance from everyday intuition, also on the level of explanation.

Still, even though mechanistic explanations are not always the most usual and natural, one could hold on to the idea that such explanations are *possible in principle*, and provide a kind of basic understanding of physical processes. As we shall discuss in section 3, it is indeed true that mechanical models can very often be constructed in physics—this was proved by Poincaré at the end of the nineteenth century. However, this existence in principle depends on a rather trivial underdetermination argument, and it is far from clear that it carries epistemological weight.

The fundamental status of mechanistic explanations was dealt a further blow by the advent of quantum mechanics in the twentieth century. It is a general trait of mechanistic explanations that they analyze the behavior of physical systems in terms of these systems' material constituents and

the interactions between them: mechanistic explanations are meant to be "decompositional" [13]. However, quantum theory has unsettled even such very general mechanistic conceptions. The very notion that a composite system can be fully analyzed in terms of the properties of its constituent parts and the relations between them has become debatable: according to standard interpretational ideas quantum theory attributes holistic features to physical systems. Thus even the most basic ingredients of the notion of a "mechanism" become moot.

Evidently, this unstable basis of the applicability of the concept of a "mechanism" has consequences for the status of mechanistic explanation in physics. Although there certainly are many cases in which mechanistic intuitions are helpful and provide understanding[1], the associated pictures and concepts cannot lay claim to fundamentality. There are fundamental physical processes in which the usefulness of mechanistic modelling is far from obvious, and in which other types of reasoning appear more fitting. On balance, mechanistic explanations cannot claim to possess a privileged status.

As we shall argue, this points in the direction of a pluralist conception of explanation and understanding in physics, according to which contextual and pragmatic factors are important in deciding which conceptual framework is the most appropriate. Mechanistic explanation is one of the tools present in the "conceptual toolbox", but depending on the specifics of the problem case other explanatory strategies may be preferable.

2 Mechanics and Mechanisms

The intuitive attractiveness and power of descriptions in terms of mechanisms, in the original and literal sense of material objects whose parts interact via pushes and pulls, is beyond dispute. We are so familiar with the operation of pulleys, drive shafts and effects of collisions, that an analysis of complicated processes in such terms provides a strong conceptual grip on what is happening. Cartesian physics lived up to exactly this ideal of mechanistic explanation, which accounts for a great deal of its contemporary popularity. However, in the Scientific Revolution Cartesian physics proved just a brief phase: its framework was insufficiently flexible for the formulation of laws of motion of the kind finally stated by Newton (in particular, as Newton argued, the law of inertia cannot adequately be stated in a

[1] We associate "understanding" with qualitative insight in the behavior of physical systems, in the sense of [6]

Cartesian material plenum without an independent non-material space-time background that serves to define what straight lines and equal time spans are).

Classical mechanics, in the form given to it by Newton, accordingly does not conform to the original mechanistic ideals. On the one hand, in addition to the traditional mechanical concepts of size, velocity and acceleration, there is a non-material arena, formed by space and time, that influences the evolution of physical processes; on the other hand, in addition to the familiar interactions via contact between impenetrable bodies action-at-a-distance forces are introduced. It is well known how Newton's contemporaries for this reason accused Newton of reintroducing mysterious occult qualities and of spoiling the progress that had been made towards clarity through a mechanistic understanding of the world.

Nevertheless, due to its enormous predictive success the new Newtonian framework before too long became the dominant scientific paradigm. A new norm for "mechanistic" became thus established: mechanistic explanation came to signify the decomposition of a material system into constituent particles, the specification of interactions via Newtonian forces between these particles, and the proof that these interactions (defined against the background of a pre-given space-time) were able to predict the observed behavior of the total system.

The forces in the original Newtonian scheme are simple "inverse square" *central* forces, i.e. forces falling off with the inverse square of the mutual distance between the interacting particles ($1/r^2$) and directed along the straight line connecting them. However, at the end of the eighteenth and the beginning of the nineteenth centuries it was found that this simple type of interaction could not account for what happens in phenomena involving magnetism and moving electrical charges, so that more complicated interaction formulas had to be written down. For example, in order to accommodate the interaction between moving electrical charges within an action-at-a-distance framework one needs forces that are not directed along the line connecting the particles and that depend not only on the particle positions but also on their velocities and accelerations. As a result, the mechanistic ideal had to be adapted once again. Instead of requiring that a process should be explained in terms of localized particles and Newtonian central forces between them, it now became sufficient to give an analysis in which the total system is decomposed into material parts with mathematically stated force laws acting between them—in addition, new properties which did not possess an immediately obvious mechanical interpretation, like electrical charge, had to be accepted.

Meanwhile, it should be recognized that the adoption of this "Newtonian" type of explanation, in terms of parts dynamically producing the whole, depends on a particular perspective on classical mechanics, which is not the only possible one. The mathematically formulated theory does not unavoidably lead to the Newtonian picture of a history unfolding in time, in which physical systems at one instant work together, via their interactions, to generate the immediate future. To start with, the idea of the "production" of new situations from the old ones as time passes does not sit well with the formalism of mechanics, or even with the structure of mathematical physics in general. This relates to the notorious problem of the "'flow of time": time occurs in mathematical physics as a parameter in the same way as the spatial coordinates, which makes it impossible to define a preferred *now*. There is no privileged point on the time axis, just as there is no preferred *here*. A fortiori, there is no definable motion of a *flow* of time, just as there is no shifting *here* within the formalism.

Of course, the notions of "here" and "now" do become applicable once an external spatio-temporal viewpoint is introduced, for example connected to an observer *who makes use of the theory*. From the internal theoretical viewpoint both the concepts of *now* and *here*, and the notion of the flow of time, are merely indexical, deriving their meaning from a reference to such an external viewpoint, and not inherent in the theory itself. This may be taken as a first indication that the choice of explanations in terms of "productive mechanisms" itself has a pragmatic and contextual background, relating to the interests of the user of the theory.

In any case, interpretations of the theory of mechanics that do not start from the assumption that the theory describes how systems change while time flows are possible, and are moreover natural when we look at the formalism from an abstract point of view. Such interpretations view the universe as laid out not only in space, but also in time, as a four-dimensional "block"—the block universe, which comprises the whole of history without making a distinction between an absolute (as opposed to an indexically defined) Past, Present and Future.

Patterns of explanation that fit in with this "static" perspective were in fact already proposed in the eighteenth and nineteenth centuries. In these alternatives to the original Newtonian approach (associated with names like Maupertuis, Euler, Lagrange and Hamilton) one does not focus on instantaneous forces that change the present physical state, but rather asks which path will be followed by a mechanical system (e.g. a particle) if it is given that it finds itself at position x_1 at instant t_1 and is located at x_2 at another time t_2. The Principle of Least Action (or more generally the Principle of

Stationary Action) states that among all possible continuous curves connecting x_1 and x_2 in the given time interval, the one actually realized minimizes (more generally: makes stationary) the "action" $S = \int_{t_1}^{t_2} L dt$.

In this formula the function L is the *Lagrangian*, defined as the difference between kinetic and potential energy of the system: $L = T - V$. The kinetic energy T ($\frac{1}{2}mv^2$ for a point particle) will involve squares of velocities, the potential energy V will usually be only a function of positions, so that the total Lagrangian $T - V$ is a function of positions and velocities.

The Principle of Least Action may suggest that a system "chooses", from all logically possible evolutions between t_1 and t_2, the one that makes the action $\int_{t_1}^{t_2} L dt$ minimal. This invites teleological patterns of explanation: the system, "knowing" that it will have to arrive at x_2 at time t_2, fulfils this task in the most economical way at its disposal.

Obviously, such anthropomorphic terminology, although not uncommon in the practice of physics, should not be taken seriously. Mathematically speaking, the Newtonian and Lagrangian approaches are equivalent: one can be derived from the other, so that arguments on the basis of the principle of least action need not introduce irreducibly novel ontological ingredients. Still, in explanations starting from the minimization of the action the focus is different than in Newtonian explanations: one now looks at the total path (if the system comprises more than one particle this path is defined in phase space), stretched out in time, and compares different possibilities. By contrast, in the Newtonian approach one focuses on the instantaneous state and computes how this state evolves in response to causal influences. This difference illustrates how the same mathematical theory (in this case classical mechanics) may be cast in various forms and how different patterns of explanation can become plausible depending on these different forms. In fact, the Lagrangian formulation is not the only non-Newtonian form that can be given to classical mechanics: the Hamiltonian and Hamilton-Jacobi formalism are still other alternatives, and there are more.

These more modern approaches in mechanics usually do not work with causal terminology (forces producing changes) and rely more on mathematical properties of the formalism. This opens up the possibility of new types of explanation, for example those based on the existence of *symmetries*. By way of illustration, if the action S (as defined above) does not explicitly depend on time (i.e. if time does not occur in the formula for S in addition to its implicit dependence on time via the coordinates and velocities—this expresses "symmetry under time translation"), it can be shown that the *energy* of the system remains constant over time (conservation of energy); if the action does not depend on position (symmetry under space translation)

it follows that *momentum* is conserved. These examples illustrate Noether's theorem, which in a general and systematic way links symmetries to the conservation of physical quantities.

Summing up, what a mechanism is, and what a mechanistic explanation amounts to, is not completely set in stone in classical mechanics. In the course of history a development into the direction of more complicated and intuitively less immediately attractive "mechanisms" has proven necessary. A constant theme in this development (until the advent of quantum mechanics, about which more in a moment) has been the notion that a system should be decomposed in its constituent particles and that the whole should be understood on the basis of the dynamics of these parts. All these various mechanistic explanations make use of intuitively plausible causal terminology (forces, production, unfolding in time) and often provide understanding. However, within the same science of classical mechanics more abstract explanations (on the basis of variational principles, symmetries, or abstract properties of the mathematical structure) are possible as well, and actually occur more frequently in advanced treatments of the subject. So explanation by "mechanisms" is not inextricably bound up with the science of mechanics: mechanics is more flexible than that, and more neutral with respect to possible patterns of explanation.

3 Maxwell's Theory and Poincaré's Theorem

The eighteenth and nineteenth centuries saw the introduction of several new types of matter, "fluids", for the purpose of explaining phenomena in a number of relatively new disciplines: chemistry, the theory of heat, and most famous and important for our topic, electrodynamics. It was mentioned in the previous section that a Newtonian treatment of moving electrical charges meets with difficulties and requires the introduction of action-at-a-distance forces of a new and unusual sort. The introduction of the electromagnetic ether and the development of the field concept, culminating in Maxwell's theory of electrodynamics of 1865, cast new light on this subject. Instead of thinking in terms of action at a distance between particles, Maxwell proposed to conceptualize the interaction between charges as mediated by undulations in an underlying medium—waves that propagated, with a finite velocity, between the charges. This new "field-theoretic" framework was a huge success: Maxwell was able to unify electricity, magnetism and optics within the same theory. As he wrote [9]: "[It seems] that light and magnetism are affections of the same substance, and that light is

an electromagnetic disturbance propagated through the field according to electromagnetic laws." Maxwell's new, general and comprehensive theory of electromagnetism made the earlier direct-action attempts obsolete.

In his 1873 *Treatise on Electricity and Magnetism* [10] Maxwell presented his definitive treatment of the theory, in which electrodynamic quantities were represented by vectors, as still usual today: for example, $\overrightarrow{\mathbf{E}}$ stands for the electric field and $\overrightarrow{\mathbf{B}}$ for the "magnetic induction". In the rapidly following further development of the discipline these vectors came to stand for locally (i.e. per spatial point) defined forces, existing throughout the medium (the ether), so that we have a continuous "field" of electric and magnetic forces $\overrightarrow{\mathbf{E}}(x)$ and $\overrightarrow{\mathbf{B}}(x)$, with x indicating position. These forces are "felt" by localized charged particles that find themselves at the position x. As derived by Lorentz (see [4]), the force exerted on a particle with electrical charge e and velocity $\overrightarrow{v}$, at position x in the field, is given by $\overrightarrow{F} = e\overrightarrow{\mathbf{E}}(x) + e\overrightarrow{v} \wedge \overrightarrow{\mathbf{B}}$ (with $\wedge$ denoting the outer product between two vectors).

The core of the theory developed in the Treatise [10] is formed by the "Maxwell equations", which govern the dynamics of the electric and magnetic fields. After a long chain of arguments, Maxwell's presentation culminates in the demonstration that this dynamics can be put in *Lagrangian form*.

The final result is a theory in which we have field quantities defined throughout space and interrelated by a set of equations (the Maxwell equations). These fields exert influences on charged particles, and charged particles in turn influence the fields; this is all described in a rigorous and abstract mathematical way. Significantly, Maxwell states that in the final analysis these field quantities express the *mechanical state* of the underlying substance, the ether. But it is not worked out how exactly this should be fleshed out: in what way does the state of motion of the material ether generate electric and magnetic fields, and how should we envisage the interaction between charged particles and the moving parts of the ether?

In 1890 Henri Poincaré published a book, *Électricité et Optique* [11], meant to explain Maxwell's theory to the French-speaking world. Poincaré starts his Introduction with the remark that for a French reader the first acquaintance with Maxwell's text will probably lead to a feeling of embarrassment and even distrust—as Poincaré states, this feeling will only disappear after much effort, and some eminent minds even keep it for ever.[2] Accord-

[2]La première fois qu'un lecteur français ouvre le livre de Maxwell, un sentiment de malaise et souvent même de défiance se mêle d'abord à son admiration. Ce n'est qu'après

ing to Poincaré there are several reasons for this reaction: the French reader wants logic, consistency and precision, preferably in the form of a deductive system with a minimum number of clearly stated axioms, and Maxwell's work does not possess this form. But there is also another reason. Behind the world of experience the French reader will wish to see another world, consisting of matter with purely geometric properties, with atoms that are point particles subjected to mechanical laws. Only then will he have the feeling that he has penetrated to the secret of the Universe.[3]

Poincaré expresses doubt about the philosophical tenability of this latter geometrical/mechanical requirement; he thinks that it concedes too much to our intuitive urge for easily visualizable pictures. Anyway, at this point in his Introduction Poincaré warns his readers that desires for an elegant axiomatic set-up and a simple and obvious mechanistic explanation will not be satisfied by Maxwell. However, and this is according to Poincaré meant to be Maxwell's most important general message to the reader of the Treatise, Maxwell *does* show that a mechanistic explanation of electric and magnetic phenomena is possible *in principle*.[4] To show that this conclusion is indeed contained in Maxwell's work, Poincaré proves a little theorem, still in the Introduction to his book [11, ix-xiv].

The proof hinges on the fact that Maxwell's theory can be given a Lagrangian formulation, as emphasized in the Treatise. Quite generally, as Poincaré is going to show, theories with a Lagrangian formulation admit a mechanical model in which masses that interact via forces derivable from a potential can be made responsible for what the theory predicts on the observable level. The idea of the proof is simple. Any physical theory should make contact with experience, and should therefore operate with physical quantities, $q_1, q_2, ..., q_n$, that are accessible to measurement. If the theory can be put in Lagrangian form, this means that there exists a potential energy function $V(q)$ of the quantities $q_1, q_2, ..., q_n$, and also a kinetic energy function $T(q, \dot{q})$ of these quantities and their time derivatives, so that the Lagrangian $L = T - V$ can be formed.

Now, if there is to be a mechanical model, it should be possible to find p

---

un commerce prolongé et aux prix de beaucoup d'efforts, que ce sentiment se dissipe. Quelques esprits éminents le conservent même toujours.

[3]Derrière la matière qu'atteignent nos sens et que l'expérience nous fait connaître, il voudra voir une autre matière, la seule véritable à ses yeux, qui n'aurait plus que des qualités purement géométriques et dont les atomes ne seront plus que des points mathématiques soumis aux seules lois de la Dynamique... C'est alors seulement qu'il sera pleinement satisfait et s'imaginera avoir pénetré le secret de l'Univers.

[4]*Maxwell ne donne pas une explication méchanique de l'électricité et du magnétisme; il se borne a démontrer que cette explication est possible* [italics in original].

values of masses, $m_1, m_2, ..., m_p$, and p positions $x_1, x_2, ..., x_p$, of p particles that together determine the measurable quantities $q_1, q_2, ..., q_n$ and in turn are functions of these measurable quantities: $x_i = x_i(q_1, q_2, ..., q_n)$.[5] The kinetic energy, when expressed in the particle quantities, should have the usual particle form: $T(q, \dot{q}) = \sum_{i=1}^{p} \frac{1}{2} m_i \dot{x}_i^2$. Since we are interested in the existence *in principle* of a mechanical model, no limit is set to the value of p, so that one may assume as many particles as one likes.

With this freedom in the number of particles it is always possible to satisfy the equation for the kinetic energy, $T(q, \dot{q}) = \sum_{i=1}^{p} \frac{1}{2} m_i \dot{x}_i^2$: the number of unknowns, p can be made much greater than the number n of known quantities. We therefore have a case of mathematical underdetermination, and there will be infinitely many possible choices for the masses and positions.

Given any such choice, we can rewrite the potential energy V as a function of the particle positions, and the Lagrangian equations of motion become: $m_i \ddot{x}_i = -\frac{dV}{dx_i}$. So we have arrived at a theoretical scheme in which p moving particles, interacting via forces $-\frac{dV}{dx_i}$, fully reproduce the empirical predictions of the theory with which we started.

In other words, under very general conditions (the existence of a Lagrangian scheme) it is possible to find *many* mechanical models that lead to the exact same predictions as the given physical theory. These models will contain a number of point masses, interacting through local forces that derive from a potential. As said, there is no lack of such models: due to the underdetermination signalled above, if there is one such model, there is an infinity of them.[6]

Seen from this perspective, mechanical explanation is always possible—but it is cheap and resembles a sleight-of-hand. It seems an empty addition to what we could already understand in terms of the quantities q alone. As Poincaré discusses further in *La Science et l'Hypothèse* [12, 196-197; 251-259], one might initially think that requirements concerning the form of the forces will give the concept of mechanical explanation more bite; for example, one could impose that the forces should be central, or expressible as fixed connections in the manner of Hertz, or perhaps reducible to the effects of direct particle collisions. But given the freedom to choose the number of particles p as large as one wishes, this will not help: the problem

[5] More precisely, each particle position has three components in three-dimensional space, so that there are four unknowns associated with each particle.

[6] As Poincaré [11, xiv] puts it: "*Si donc un phénomène comporte une explication méchanique complète, il en comportera une infinité d'autres qui rendront également bien compte de toutes les particularités révélées par l'expérience.*" [italics in original]

will remain underdetermined and there will still be infinitely many solutions.

Poincaré concludes that the choice for a mechanistic explanation will necessarily involve non-empirical, personal and pragmatic factors. He suggests that future physicists will no longer be interested in thinking about such things and will leave this to metaphysicians; and that in the end the reader of Maxwell's Treatise will see the artificial elements in the [mechanical] theoretical schemes that he once admired.[7]

4 Mechanisms and Quantum Mechanics

A common theme in the various forms of mechanistic explanation that we have considered is that they are decompositional [13]: the behavior of a composite system is explained by reference to its material parts and the interactions between these parts. This has become the motivation for the "New Mechanicism" in the philosophy of science. This New Mechanicism is meant to be an elaboration of and improvement on Salmon's causal scheme of explanation, according to which good explanations are ontologically grounded in the objectively existing causal structure of the world [14]. The new mechanicists share this "ontic" commitment, but work out the details of the causal structure in a way that is slightly different from Salmon's original proposals, namely in terms of *mechanisms*, defined as complex, composite systems whose efficacy in performing a certain function can be understood on the basis of the concerted action of its parts. Glennan gives the following definition [7]:

> a mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations.

This definition accords well with the nature of the mechanisms that passed review in our historical sketch of classical mechanics. In particular, the behavior to be explained is *produced* by the *interactions* between the *parts*; and as Glennan explains, these parts must be *objects* with a high degree of robustness or stability, which are generally spatially localized. The interactions bring about changes in the properties of one part as a consequence of

[7] Un jour viendra peut-être où les physiciens se désintéresseront de ces questions, inaccessibles aux méthodes positives, et les abandonneront aux métaphysiciens. ... Le lecteur...finit par comprendre ce qu'il y avait souvent d'un peu artificiel dans les ensembles théoriques qu'il admirait autrefois [12, 258-259].

changes in the properties of another [7, S344]. As Glennan adds concerning these interactions [7, S352], "events occurring at some point in space and time are explained as the consequence of the operation of causal mechanisms operating in that region of space and time. Our global evidence suggests that—quantum mechanics aside—causality is everywhere local."

The picture is that explanation by mechanisms is ontologically privileged as it latches on to the objective structure of the world: the world consists of composite objects whose behavior is produced by local interactions between localized component systems. It is important for this new mechanicism, as it was for older forms of mechanicism, that the interaction between any two components should be the same as in the case in which these components are the only systems present: the interactions should not be "holistic", depending on the behaviour of the complex system that is to be explained. The mechanistic intuition is that the global system should be reducible to its parts. No wonder then that Glennan added the clause "quantum mechanics aside" in the just-quoted statement: quantum mechanics is notorious for the problems it engenders for practically all of the mentioned ingredients of mechanistic explanation: according to quantum mechanics, physical systems need not be localized, interactions possess non-local aspects, and perhaps most important of all, the properties of a composite system can generally not be reduced to the properties of its parts.

There is one underlying reason for all these problems. Von Neumann already pointed out, in his seminal 1932 book on the mathematical structure of quantum mechanics [15], that *the* central novel feature of quantum theory is that states of physical systems are to be represented by *vectors* in a state space (a Hilbert space), with the property that the *superposition* (sum) of any two such vector states again represents a realizable state. Accordingly, the structure of the quantum state space is radically different from what we are used to in classical physics. For example, if we have two vector quantum states denoted by $|x_1\rangle$ and $|x_2\rangle$, meant to refer to a system at position x_1 and x_2, respectively, the sum state $\frac{1}{\sqrt{2}}(|x_1\rangle + |x_2\rangle)$ is again a *bona-fide* state—but this time we have a state that does not correspond to one definite localization. The superposition principle, saying that any two states may be superposed to form a new state in which a physical system can find itself, is responsible for most non-classical features of quantum mechanics.

In particular, the superposition principle explains why the state of a composite quantum system generally cannot be reconstructed from the states of its component parts. Suppose that we have a system C that consists of the two partial systems A and B; and suppose that possible states of A and

B are ($\{|\alpha\rangle_i\}$ and $\{|\beta\rangle_i\}$ (these are vectors in the Hibert spaces associated with A and B, respectively). In this situation, a simple composite state of C is

$$|\Psi\rangle = |\alpha\rangle_k \otimes |\beta\rangle_k, \tag{1}$$

which can be interpreted in a classical way: The system C (represented by $|\Psi\rangle$, a vector in the Hilbert space associated with C) consists of two components, A and B, with states $|\alpha\rangle_k$ and $|\beta\rangle_k$, respectively, and the properties of C supervene on these of A and B. The crucial point is that the superposition principle tells us that a superposition of states of the form (1) is also possible, which leads to a state of the form:

$$|\Psi\rangle = \sum_i c_i |\alpha\rangle_i \otimes |\beta\rangle_i, \tag{2}$$

where $|\alpha\rangle_i$ and $|\beta\rangle_i$ are again state vectors in the Hilbert spaces of A and B, respectively, and the coefficients c_i are complex numbers. In the situation represented by Eq. 2 the global system C is in a so-called pure state (represented by a vector in Hilbert space), but the partial systems A and B, taken by themselves, are in "mixed states". This can be intuitively understood from Eq. 2 because both A and B are associated with a whole range of state vectors ($\{|\alpha\rangle_i\}$ and $\{|\beta\rangle_i\}$, respectively) so that it is not implausible that they are best represented by a mixture of these states. Now, as von Neumann showed, it is a mathematical fact that $|\Psi\rangle$ fully determines the mixed states of A and B, but that the reverse is not true: The mixed states of the component systems, in a situation of the type represented by Eq. 2, do not fix the state of C, i.e. $|\Psi\rangle$. It follows from this that knowledge of all physical properties of A and B individually, and all possible outcomes of measurements performed on A and B by themselves, does not suffice to determine the state of C. There thus exists a certain holism in quantum mechanics: properties of a whole do generally not supervene on properties of the parts.

The answer to the question of whether one can think of composite quantum systems as being built up from parts that interact via local interactions (a "local model") relates to the just-sketched holism. As famously proved by Bell [2], it is impossible in certain total states of the form (2) to reproduce, with a local model, the quantum mechanical predictions for the correlations between outcomes of measurements performed on A and B separately. The relevant quantum mechanical predictions have been impressively confirmed in many experiments, so that the conclusion is justified that nature is not correctly described by models with local interactions between parts—which clearly are *mechanisms* of the sort discussed earlier.

We already mentioned the general lack of localizability of quantum systems, which leads to another discrepancy between the quantum ontology and the ontology of local mechanisms. Since quantum states may be superpositions of states that are (more or less) localized, the resulting states can have very extended spatial domains. This is important for practical applications, as illustrated by the famous "double-slit" example, in which a single electron goes to two slits at the same time, in spite of a substantial distance between the slits. In the beginning days of quantum theory examples like this were mere thought experiments, but now they are routinely realized in laboratories and prove to be important for practical applications.[8]

The quantum world is therefore strange and radically non-classical, as emphasized in many accounts of the theory. Nevertheless, it is clear that if the theory is to be empirically adequate, classical patterns of behavior have to emerge in some situations—after all, there must be a reason that classical mechanics was successful for so long a time. If there were no classical limiting situations, classical physics would never have developed. The details of the classical limit of quantum mechanics are to some extent controversial, because they relate to interpretational issues (in particular, the measurement problem). However, there is a growing consensus that the process of "decoherence" is of vital importance here.

Decoherence occurs when a quantum system couples to its environment—usually an environment with very many degrees of freedom, which makes the process practically irreversible. The interaction with this environment is governed by the usual quantum mechanical evolution (the Schrödinger equation or a relativistic generalization of it); it is a case of ordinary quantum mechanical interaction. As we shall see in a moment, one of the effects of decoherence is that the effects of entanglement and superposition become hard to detect.

An entangled state has the general form shown in Eq. 2. Now suppose that there is an environment E that interacts with the system in the state $|\Psi\rangle$ of Eq. 2, such that E responds differently to the different terms in (2).

[8] We here follow standard interpretational ideas, staying close to the standard Hilbert space formalism. The interpretation of quantum mechanics is notoriously controversial, and there are proposals that differ from the standard account. The difficulties mentioned in the text assume different forms depending on the interpretation that is being considered, but in any interpretation there remain holistic and non-local features. For example, in the Bohm version of quantum mechanics [3] particles *are* localized, but they interact via non-local forces of a holistic sort: the form of these action-at-a-distance forces depends on the quantum state of the composite object. So also here there is no supervenience of the whole on the parts.

This can be represented mathematically by the following evolution:

$$|\Psi\rangle|E_0\rangle = \sum_i c_i|\alpha\rangle_i \otimes |\beta\rangle_i|E_0\rangle \mapsto \sum_i c_i|\alpha\rangle_i \otimes |\beta\rangle_i|E_i\rangle. \qquad (3)$$

In this formula the symbol $\mapsto$ represents the evolution: this evolution maps the initial state on the left hand side of the symbol into the final state on the right hand side. $|E_0\rangle$ is the initial environment state; the states $|E_i\rangle$ are the environment states that couple to the states $|\alpha\rangle_i \otimes |\beta\rangle_i$ of the composite object.

The crucial fact that makes decoherence so important is the following. When one performs a measurement on the composite system alone, after its interaction with the environment E, the typical effects of entanglement will be blurred. In the extreme case that the states $|E_i\rangle$ are mutually *orthogonal* (i.e. the states are without any overlap—this is the case if the environment responds completely differently to the various states $|\alpha\rangle_i \otimes |\beta\rangle_i$) the effects of entanglement will even become completely invisible in measurements on the composite system alone (i.e. if one does not look at E).[9]

It should be noted, however, that this disappearance of entanglement is not only approximate, but also relative to a limited class of observations. As inspection of Eq. 3 demonstrates, the total state of the original composite system plus its environment is still entangled—the process of decoherence has merely spread out the original entanglement so that it now also involves the environment E. As a consequence, observations of the original system *plus* the environment with which it has interacted will show the entangled nature of the total state, with its non-classical and non-local characteristics. However, it is true that if one restricts oneself to measurements on open systems *without* looking at their environments, *and* if one's measurements are not too precise, quantum effects will often[10] not manifest themselves and classical models of what happens will become possible. Another important consequence of decoherence is that open quantum systems in environments

[9]Formally, the essential difference between the situations before and after interaction with the environment is that initially the expectation value of any operator O of the composite system alone is given by $\langle\Psi|O|\Psi\rangle = \sum_{i,j} c_i^* c_j \langle\alpha_i\beta_i|O|\alpha_j\beta_j\rangle$, whereas after the interaction with the environment this becomes $\sum_{i,j} c_i^* c_j \langle\alpha_i\beta_i|O|\alpha_j\beta_j\rangle\langle E_i|E_j\rangle$. The inner products $\langle E_i|E_j\rangle$ that have appeared tend to wash out the "cross terms", with $i \neq j$—these cross terms are needed to show the presence of entanglement. In the extreme case of orthogonality between different environment states we have $\langle E_i|E_j\rangle = 0$ if $i \neq j$, so that the effects of entanglement vanish completely from sight.

[10]In particular, in the circumstances of everyday observation. In laboratory experiments it turns out that quantum effects affecting even macroscopic objects can be made visible—these experiments on so-called Schrödinger cat states have become almost routine now.

of the kind we are used to tend to become *localized*. This is because the usual interactions (electromagnetism, gravity) are sensitive to position, with the consequence that (practically) orthogonal environment states will become correlated to object states associated with different positions. By virtue of the same argument as before, superpositions of different positions will therefore become practically unobservable.

Summing up, quantum mechanics describes a world that is basically non-local and holistic, with properties of composite systems that generally do not supervene on the properties of their parts. But the process of decoherence is able to hide these typical quantum features from view. In particular, when we make observations on Earth, outside a fundamental physics laboratory, models on the basis of classical physics will normally work very well.

So we may conclude that quantum mechanics leaves room for mechanistic explanations: there is a limited domain of quantum phenomena, defined by a) a restriction on which parts of a total system are investigated, b) a limitation on the accuracy with which these investigations are carried out, and c) the presence of decoherence processes, in which mechanistic models apply. This seems in accordance with a conclusion recently drawn by Kuhlmann and Glennan, who write [8, 353]

> that decoherence provides a useful explanation of why, in particular local circumstances, systems behave classically in spite of their being ultimately constituted of entities that obey the principles of quantum mechanics, and that this explanation deflects possible concerns over the ontological and explanatory legitimacy of the mechanistic approach.

However, one should not overrate this result.[11] Although mechanistic models usually make extremely accurate predictions in familiar "classical" settings, taken completely literally these predictions are, even though very close, still wrong: a purely quantum mechanical calculation, taking into account entanglement and the non completely vanishing values of the factors $\langle E_i|E_j\rangle$ (see note 9) will give other and, importantly, *better* predictions. So there are features of reality, detectable in principle, that show that the literal content of the ontological claims of the mechanistic explanation strategy is *false*.

[11] Kuhlmann and Glennan sometimes make statements that create the (what would be a mistaken) impression that there is absolutely nothing wrong with mechanistic explanations in semi-classical contexts, even given the validity of quantum mechanics; e.g. they say "In this paper we argue, in part by appeal to the theory of quantum decoherence, that the universal validity of quantum mechanics does not undermine neo-mechanistic ontological and explanatory claims as they occur within classical domains" [8, 337].

The situation can be compared to others that we already encountered. Mechanical explanation by central forces, even though it was for some time the paradigm of mechanical explanation, turned out to be empirically inadequate when electrodynamic phenomena were investigated more extensively and in more detail. The law of Coulomb, according to which two electrical charges attract or repel each other by an inverse-square central force (in analogy to Newton's law of gravity) had to be replaced by a more complicated law for moving charges (this more complicated interaction can be derived from Maxwell's equations). Now, in many circumstances—in particular those that were familiar to researchers in the beginning of the nineteenth century—Coulomb's law still yields excellent predictions in spite of this complication. This is because the charges under investigation often do not move too fast—although they always move somewhat and are never perfectly stationary—and the deviations from Coulomb's law are minute anyway, hardly observable without ultra-sensitive experimental techniques. So there is a certain domain of electrodynamic phenomena that in spite of the validity of Maxwell's theory can be handled perfectly well, for all practical purposes, with the older Coulomb theory. Does this justify the conclusion that Maxwell's electrodynamics does not undermine the older ontological and explanatory claims? It seems clear that this is not the case. True, *explanations* by means of the Coulomb theory can often still be maintained after Maxwell, but these explanations depend for their ontological grounding on Maxwell's electrodynamics plus an argument that the new dynamical effects that occur (e.g., loss of energy by radiation) fall below the threshold of observational accuracy. Similar comments apply to many other examples from the practice of physics, in which explanations are still given on the basis of obsolete and false theories.[12] In all these cases the original ontological basis of the explanations *is* undermined, but this does not exclude that the explanations themselves, as argumentative patterns, are still useful.

[12]It is sometimes argued that it is impossible to give valid explanations on the basis of false theories at all (see [5], and the volume of which that essay is a chapter, for recent discussions on this topic). We do not agree that it is impossible to explain without literal truth (see the next section)—but if this impossibility were to be accepted, this would clearly call the mechanist ideal into question in a more radical way than we do here.

5 Conclusion: Mechanisms, Explanations and Understanding in Physics

Even within classical mechanics the status of mechanistic explanation is not unchallenged. It is true that in the days when the theory was first proposed the new ideal according to which all physical processes should be explained as the result of pushes and pulls sparked off great enthusiasm, but this ideal had soon to be abandoned. The rules of mechanicism had to be stretched, first by admitting central action-at-a-distance forces, then by allowing the forces to become more complicated. The reason, of course, was the development of physical theory: theoretical schemes along Cartesian lines proved to be empirically inadequate, after some time central forces shared this fate, and not long thereafter the whole concept of action-at-a-distance forces became obsolete. In the meantime new mathematical frameworks had developed for the formulation of classical mechanics, like the Lagrangian an Hamiltonian formalisms, and these gave rise to very different patterns of explanation, for example via variational principles.

Still, even though the Cartesian push and pull paradigm has long been left behind as a fundamental and general scientific scheme, explanations along these lines remain useful. For example, even if we think that interactions between bodies are always mediated by fields (perhaps quantum fields) or complicated action-at-a-distance forces, it usually helps to visualize such interactions via the picture of Cartesian collisions—physics textbooks are full of pictures of this kind, even if the subject is quantum field theory. This is because simple mechanistic models, if they yield results that are not too far off the mark, provide us with qualitative understanding of a process: they enable us to see, without entering into detailed calculations, what the approximate outcome of a process will be. The familiarity of the push and pull scheme makes it intuitively manageable.

The same comments apply to the other types of mechanistic explanation. For example, even in the context of general relativity it often helps to think of the gravitational attraction between material bodies in terms of the conceptual framework of Newton's theory. Why is it that light cannot escape from a black hole? Because the black hole is so massive that the gravitational force it exerts on light pulls the light towards the black hole so strongly that it cannot get away. Is is easy to understand the process this way, and it requires a lot more training to become equally familiar with the general relativistic scheme of null-geodesics, horizons and the Einstein field equations. Nevertheless, it is not impossible at all to acquire an intuitive

familiarity with such advanced mathematical schemes; seasoned researchers do not need to make detailed calculations in order to make a qualitative judgment about, e.g., to what extent a solution of the field equations will deviate from Euclidean geometry, given a particular mass distribution.

The Lagrangian approach to classical mechanics, with its variational principle, illustrates our general point further. Although this approach pertains to cases in which Newtonian explanations that use forces are certainly also possible, there are circumstances in which one may nevertheless prefer an explanation along Lagrangian lines. This may happen if one wants to avoid anthropomorphic or indexical elements in one's explanations (see the discussion about the flow of time and four-dimensionality in section 2), or if one wishes to lay stress on the continuity between classical mechanics and relativity theory; or on the analogies between mechanics and optics. Explanations via the Lagrangian framework are certainly not less ontologically grounded than their Newtonian causal counterparts: the structure of the four-dimensional world is such that it obeys variational principles—if anything, it is the Newtonian mechanistic explanation, with its *production* of effects during *the flow of time* that can be accused of introducing subjective elements. Moreover, in many cases one can develop an intuitive feeling for the outcome of variational arguments so that they make it possible to achieve understanding. For example, it is understandable why the trajectory of a free particle will be a straight line, as this path realizes the shortest distance. Also in cases with more complicated Lagrangians a similar geometric interpretation often makes it easy to make qualitative statements about the form of trajectories as shortest connection in some geometry.

We know from section 3 that mechanistic explanations will be available in principle as soon as a Lagrangian scheme applies—for this, the Lagrangian does not even need to depend on mechanical quantities at all. What is more, there will be infinitely many different mechanistic explanations, of any sort one wishes: using contact forces, action at a distance, etc. It hardly needs argumentation, though, that this abundance does not help to enhance the attractiveness of such explanations: they will as a rule be too unwieldy, complicated and unnatural to be taken seriously. Clearly, the mere fact that the explanations in question are mechanistic does not compensate for this disadvantage—such "Poincaré schemes" are artificial and unenlightening, even though they are able to reproduce all empirical results correctly.

What this all points to is that explanation and understanding in physics are not restricted to one privileged standard format. There are usually several forms of explanation available, and which type is actually chosen in a particular situation depends on contextual factors like the exact question

that is being asked (and its "contrast class"), the intended use of the explanation, and the conceptual framework that is adopted. The same applies to the notion of *understanding*. Since understanding is a more qualitative concept than explanation and also depends on factors like the skill of the actor who is involved and his/her familiarity with the theoretical framework, there is even more freedom here than in the case of explanation [6, 5]. In particular, it is not unusual to achieve understanding of physical processes with the help of obsolete theories that have been proven wrong when taken literally. Such theories may in spite of their incorrectness provide tools by which one can attain an intuitive grip on a process, and succeed in foreseeing its outcome in a qualitative way. This is particularly true for theories that use mechanical concepts. Quantum mechanics has supplanted these theories, but mechanical reasoning may still provide a conceptual grip on certain phenomena.

A final example may be useful here. In their plea that the validity of mechanistic explanation is not undermined by quantum mechanics, Kuhlmann and Glennan write [8, 357]

> why do flocks of birds so often form the inverted-l-shaped form often seen in autumn? A mechanistic explanation explains how this local phenomenon arises through the local interaction of the birds; global entanglement between the birds (and their constituents) and the rest of the universe are (to a high approximation) not causally or explanatorily relevant to the production of this phenomenon.

This is exactly right if construed as a proposal for one way of *understanding* how the shape of a flock of birds arises. But note that a mechanistic explanation in terms of productive forces is not the only possibility of achieving such understanding: a Lagrangian variational approach (e.g., in this case, in terms of finding a constellation of birds with maximum stability, by minimizing an energy expression) would work as well; it depends on contextual factors which approach is preferred.

Note further that the mechanical theory invoked in the explanation of the form of the birds flock is, taken literally, *wrong* (as acknowledged in the quotation by the addition of "to a high approximation"). Quantum theory is taken to be the more correct theory here, and it is in fundamental conflict with classical mechanics. It follows that a *better* explanation than the suggested mechanistic one is available if one is interested in the highest attainable predictive accuracy. It is true that the differences in cases like this will normally be astronomically small, but still in principle the suggested

mechanical model will give results that are wrong in its details. The general statement that the quantum aspects of the situation are not causally or explanatorily relevant is therefore false. It depends on the *perspective* that is taken whether or not entanglement and other quantum aspects should be taken into account. From an already taken mechanical vantage point they do not play a role; from the perspective of quantum mechanics they *are* relevant. In particular, these very quantum aspects determine whether an approximate account in mechanistic terms will be viable at all; they are thus certainly relevant in an explanatory sense.

We therefore conclude that on the fundamental ontological level physics has moved away from mechanisms: quantum mechanics is fundamentally at odds with the image of composite systems whose properties are produced by the properties of their parts, via local interactions. In spite of this, for the purposes of explanation and understanding mechanistic reasoning remains an important conceptual tool. But it is not at all the only possible one: a toolkit of conceptual instruments is available, and it depends on contextual factors which one should be chosen.

References

[1] Dijksterhuis, E.J. (1961). *The Mechanization of the World Picture*. New York: Oxford University press.

[2] Bell, J. (1964). On the Einstein Podolsky Rosen Paradox. *Physics*, 1, 195–200.

[3] Bohm, D. (1952). A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables, I, II. *Physical Review*, 85, 166–179; 180–193.

[4] Darrigol, O. (2000). *Electrodynamics from Ampère to Einstein*. Oxford: Oxford University Press.

[5] De Regt, H.W. and Gijsbers, V. (2016). How False Theories Can Yield Genuine Understanding. Chapter 3 in S. Grimm, C. Baumberger and S. Ammon (eds.) *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*. New York: Routledge.

[6] De Regt, H.W. and Dieks, D. (2005). A Contextual Approach to Scientific Understanding. *Synthese* 144, 137–170.

[7] Glennan, S. (2002). Rethinking Mechanistic Explanation. *Philosophy of Science*, 69, Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association. Part II: Symposia Papers, S342–S353.

[8] Kuhlmann, M. and Glennan, S. (2014). On the Relation between Quantum Mechanical and Neo-Mechanistic Ontologies and Explanatory Strategies. *European Journal for Philosophy of Science*, 4, 337–359.

[9] Maxwell, J. C. (1865). A Dynamical Theory of the Electromagnetic Field. *Philosophical Transactions of the Royal Society of London*, 155, 459–512.

[10] Maxwell, J. C. (1873). *A Treatise on Electricity and Magnetism*, Volumes I, II. Oxford : Clarendon Press.

[11] Poincaré, H. (1890). *Électricité et Optique*. Paris: G. Carreé.

[12] Poincaré, H. (1902). *La Science et l'Hypothèse*. Paris: E. Flammarion. English translation (1905): *Science and Hypothesis*. London: Walter Scott Publishing Company.

[13] Psillos, S. (2011). The Idea of Mechanism. In P.M.K. Illari, F. Russo and J. Williamson (eds.), *Causality in the Sciences*, 771–788. Oxford: Oxford University Press.

[14] Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

[15] von Neumann, J. (1932). *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer. English translation by Robert T. Beyer (1955): *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press.